Memory Load Alters Perception-Related Neural Oscillations during Multisensory Integration

Georgios Michail,¹ Daniel Senkowski,¹ Michael Niedeggen,² and Julian Keil³

¹Department of Psychiatry and Psychotherapy, St. Hedwig Hospital, Charité–Universitätsmedizin Berlin, Berlin 10115, Germany, ²Department of Education and Psychology, Free University Berlin, Berlin 14195, Germany, and ³Biological Psychology, Christian-Albrechts-University Kiel, Kiel 24118, Germany

Integrating information across different senses is a central feature of human perception. Previous research suggests that multisensory integration is shaped by a context-dependent and largely adaptive interplay between stimulus-driven bottom-up and top-down endogenous influences. One critical question concerns the extent to which this interplay is sensitive to the amount of available cognitive resources. In the present study, we investigated the influence of limited cognitive resources on audiovisual integration by measuring high-density electroencephalography (EEG) in healthy participants performing the sound-induced flash illusion (SIFI) and a verbal *n*-back task (0-back, low load and 2-back, high load) in a dual-task design. In the SIFI, the integration of a flash with two rapid beeps can induce the illusory perception of two flashes. We found that high compared with low load increased illusion susceptibility and modulated neural oscillations underlying illusion-related cross-modal interactions. Illusion perception under high load was associated with reduced early β power (18–26 Hz, ~70 ms) in auditory and motor areas, presumably reflecting an early mismatch signal and subsequent top-down influences including increased frontal θ power (7–9 Hz, ~120 ms) in mid-anterior cingulate cortex (ACC) and a later β power suppression (13–22 Hz, ~350 ms) in prefrontal and auditory cortex. Our study demonstrates that integrative crossmodal interactions underlying the SIFI are sensitive to the amount of available cognitive resources and that multisensory integration engages top-down θ and β oscillations when cognitive resources are scarce.

Key words: β oscillations; multisensory integration; sound-induced flash illusion; θ oscillations; top-down; working memory

Significance Statement

The integration of information across multiple senses, a remarkable ability of our perceptual system, is influenced by multiple context-related factors, the role of which is highly debated. It is, for instance, poorly understood how available cognitive resources influence crossmodal interactions during multisensory integration. We addressed this question using the sound-induced flash illusion (SIFI), a phenomenon in which the integration of two rapid beeps together with a flash induces the illusion of a second flash. Replicating our previous work, we demonstrate that depletion of cognitive resources through a working memory (WM) task increases the perception of the illusion. With respect to the underlying neural processes, we show that when available resources are limited, multisensory integration engages top-down θ and β oscillations.

Introduction

The ability to integrate information across multiple senses is a fundamental aspect of our perceptual system. Multisensory

integration is subject to both bottom-up stimulus-driven and top-down endogenous influences (Talsma et al., 2010). Moreover, the relative contribution of these influences is highly adaptive and depends on various stimulus-related and task-related parameters (Welch and Warren, 1980; van Atteveldt et al., 2014). An interesting open question is whether working memory (WM) load affects the equilibrium between bottom-up and top-down influences during multisensory integration (Macaluso et al., 2016; Michail and Keil, 2018). To study the effect of WM load on multisensory integration, as expressed in multisensory perception and neural oscillations, we used the sound-induced flash illusion (SIFI) paradigm (Shams et al., 2002; Keil, 2020). In the SIFI, an audiovisual stimulus comprising one flash and two rapid beeps can evoke either the perception of one

Received June 2, 2020; revised Oct. 2, 2020; accepted Oct. 29, 2020.

Author contributions: G.M. and J.K. designed research; G.M. performed research; G.M. analyzed data; G.M., D.S., M.N., and J.K. wrote the paper.

This work was supported by the German Research Foundation Grant KE1828/4-1 (to J.K.) and by a Charité–Universitätsmedizin Berlin scholarship (G.M.). We thank Teresa Ramme, Amal Sarhan, and Alex Masurovsky for their assistance in the data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Georgios Michail at georgios.michail@charite.de.

<https://doi.org/10.1523/JNEUROSCI.1397-20.2020>

Copyright © 2021 the authors

(no illusion) or two flashes (illusion). Contrasting brain responses to audiovisual stimuli that evoke these two perceptual states allows the investigation of crossmodal interactions underlying the SIFI (Keil et al., 2014; Kaiser et al., 2019).

Research suggests that neural oscillations can orchestrate multisensory processing and that different frequency bands reflect the involvement of bottom-up and top-down influences (Keil and Senkowski, 2018). The bottom-up integration of simple audiovisual stimuli has been linked to γ band oscillations (Senkowski et al., 2005, 2007). Similarly, the perception of SIFI has been associated with γ oscillations in sensory and higher order association areas (Mishra et al., 2007; Balz et al., 2016a,b). This suggests that the SIFI relies on bottom-up crossmodal interactions. In contrast, integration of audiovisual speech stimuli in the McGurk effect (McGurk and MacDonald, 1976), where an illusory speech phoneme is perceived upon presentation of an auditory phoneme together with incongruent visual lip movements, relies not only on γ oscillations (Kaiser et al., 2005), but also on top-down crossmodal interactions mediated by frontal θ oscillations (Keil et al., 2012; Roa Romero et al., 2016; Fernández et al., 2018) and fronto-central β oscillations (Roa Romero et al., 2015; Kumar et al., 2016). The reliance of SIFI on bottom-up interactions renders SIFI an optimal paradigm to examine the impact of orthogonal manipulations of top-down influences, such as WM load, on multisensory integration. The central question therein is whether the perceptual integration of audiovisual signals in the SIFI and the underlying bottom-up interactions are sensitive to the depletion of available resources by WM load.

To address this question, we recorded electroencephalography (EEG) in participants performing a dual-task paradigm comprising the SIFI combined with an orthogonal *n*-back task, which was used to manipulate WM load. We expected that higher WM load would lead to increased SIFI illusion rates (Michail and Keil, 2018). In addition, we hypothesized that neural oscillations underlying the SIFI would be affected by the depletion of cognitive resources under high load. A theoretical framework proposed that top-down attentional control is needed when competition between the unimodal constituents of a multisensory stimulus is high (Talsma et al., 2010). For this reason, we anticipated that increased WM load would reduce the available resources, thereby increasing the competition between auditory and visual input in the SIFI task and leading to recruitment of top-down mechanisms during the SIFI. Based on the central role of frontal θ oscillations in cognitive control (Cavanagh and Frank, 2014), and the involvement of β oscillations in mediating top-down influences (Arnal and Giraud, 2012; Fries, 2015), we hypothesized that WM load would be associated with modulations in frontal θ and β power during multisensory integration.

Materials and Methods

Participants

Forty participants (mean age $\pm$ SD: 26.6 ± 7.8 years; 19 females) with normal hearing, normal or corrected-to-normal vision and no history of neurologic disorders were recruited for the study. Previous studies have demonstrated a large interindividual variability in the perception of SIFI (Mishra et al., 2007; Keil et al., 2014; Hirst et al., 2020). Selection criteria were defined in line with previous research (Michail and Keil, 2018; Kaiser et al., 2019). Eight participants were excluded from further analyses because they did not perceive the SIFI illusion in at least one of the different load conditions [illusory percept in $<10\%$ ($n=4$) or $>90\%$

($n=4$) of all critical A_2V_1 trials]. Eight additional participants with $<60\%$ correct trials in one of the control conditions (see below, sections Stimuli and SIFI) were also excluded. Therefore, a subset of 24 participants (mean age $\pm$ SD: 26.6 ± 7.4 years; 13 females) was selected for the behavioral analysis. Three additional participants with excessive EEG artefacts (slow wave drifts and muscular artefacts) were excluded during EEG preprocessing. Therefore, a subset of 21 participants (mean age $\pm$ SD: 25.9 ± 7.3 years; 13 females) was entered into further EEG analyses. All participants provided written informed consent. The study was conducted in accordance with the 2008 Declaration of Helsinki and approved by the ethics committee of the Charité–Universitätsmedizin Berlin (approval number: EA1/207/15).

Stimuli

Stimulus presentation and recording of participants' responses were implemented using the Psychophysics toolbox (Brainard, 1997; RRID:SCR_002881) for MATLAB (The MathWorks). The study was conducted in a dimly lit, electrically shielded, noise-attenuating chamber. Visual stimuli were displayed on a 21-inch CRT screen at a distance of 1.2 m with a 75-Hz refresh rate. Auditory stimuli were controlled by a USB audio interface (UR22mkII, Steinberg) and delivered through in-ear headphones (ER30, Etymotic Research).

n-back task

Stimuli for this task were upper case letters, that were presented in white color on a gray background at the center of the screen. For each block, a pseudorandom sequence of letters was selected from the set of English consonants. To avoid the use of phonemes as a strategy, vowels were excluded. In 0-back trials the target was always the letter X. To ensure equal task difficulty in all the 2-back sequences, we explicitly manipulated the sequences to exclude the occurrence of lure trials. Lure trials are potentially confusing because in these trials the presented letter is the same as the one presented in the previous trial. Thirty-three percent of 0-back and 2-back trials were targets.

SIFI

Six stimulus combinations were presented, consisting of 0, 1, or 2 auditory (A) stimuli combined with either 0, 1, or 2 visual (V) stimuli (A_0V_1 , A_0V_2 , A_1V_1 , A_2V_0 , A_2V_1 , A_2V_2). The visual (flash) stimulus was a white disk subtending a visual angle of 1.6° and was presented at 4.1° centrally below the fixation cross, for 13.33 ms. The auditory (beep) stimulus was a 78-dB (SPL) 1000-Hz sine wave tone that was presented for 7 ms.

Experimental design

Participants performed a dual task paradigm (Fig. 1A), which combined a visual verbal *n*-back task and the SIFI paradigm (Shams et al., 2002; Keil, 2020). The experiment included 888 trials and was divided in 12 blocks (6 blocks for each load level: 0-back and 2-back). The order of blocks was randomized across participants. Each block consisted of 74 trials and included 34 critical A_2V_1 trials [which induce the percept of either one (no illusion) or two flashes (illusion)] and 40 control trials, i.e., right trials for each of the other five audiovisual combinations. Including breaks, the duration of the experiment was ~ 80 min. Participants performed 10 practice trials for each load condition before the start of the experiment. Each trial started with a central fixation cross, which was presented for 500 ms (Fig. 1B). Then, a letter was presented for 500 ms, followed by the display of fixation cross for 1500 ms, during which participants were asked to indicate, by a button press, if the presented letter matched the letter X (0-back, low load) or the letter presented two trials before (2-back, high load). No response was required for non-targets. Following the 1500 ms response window, a fixation cross was displayed for a variable duration of 500–800 ms, followed by the presentation of one of the six SIFI stimulus combinations. In combinations including two auditory or two visual stimuli the stimulus-onset asynchrony (SOA) was 53.3 ms. After the presentation of a stimulus from the SIFI task, the fixation cross was displayed again and participants had to indicate the number of perceived flashes by a button press (three buttons: 0, 1, or 2). Following the button press or after 1500 ms (if no button was pressed), a

Figure 1. Illustration of the dual-task paradigm. **A**, In the first part of each trial (*n*-back task) participants had to indicate whether the letter is a target ('X' in the 0-back condition and same letter as the one presented two trials before in the 2-back condition). In the second part of each trial, an audiovisual stimulus of the SIFI task was presented, and participants reported the number of flashes they perceived. **B**, Overview of structure and timing of a single trial. In this example, the presentation of an *n*-back letter was followed by the A_2V_1 SIFI stimuli. In the critical A_2V_1 trials, participants typically perceive one flash (no illusion) or two flashes (illusion).

new trial started. Participants reported the response for the *n*-back task with the right index finger and the number of flashes with the right thumb using a handheld gamepad (Logitech Gamepad F310, Logitech).

Data analysis

Behavioral data analysis

The *n*-back performance was assessed in terms of the sensitivity *d* prime index (*d'*) and reaction times (RTs). The *d'* takes into account both the hit rate (i.e., correctly identified targets) and the false alarm rate (i.e., wrong responses when non-target letters were presented) and was calculated using the formula $d' = Z_{\text{hit rate}} - Z_{\text{false alarm rate}}$, where *Z* is the inverse of the cumulative Gaussian distribution (Haatveit et al., 2010). Higher *d'* values indicate better *n*-back performance. Regarding the SIFI task, the RTs as well as the percentage of trials in which participants reported 0, 1, or 2 perceived flashes were estimated for each audiovisual stimulus combination. The susceptibility to the SIFI illusion (or "illusion rate") was quantified as the percentage of A_2V_1 trials in which participants reported two flashes. Paired-samples *t* tests were used to compare behavioral parameters between experimental conditions and Bayes factors (BFs) were estimated as measures of relative evidence (Rouder et al., 2009). A BF smaller than 0.33 indicates that there is evidence supporting the null hypothesis (H_0), while a BF > 3 indicates support for the alternative hypothesis (H_1). Additionally, Spearman's rank correlation analyses were performed to investigate the relationship between load-dependent (2-back minus 0-back) changes in *n*-back performance and the SIFI illusion rate. Holm–Bonferroni correction (Holm, 1979) was used when needed to account for multiple testing.

EEG recording and preprocessing

EEG was recorded using a 128-channel passive system (EasyCap) at a sampling rate of 2500 Hz. Two electrodes, at the right lateral canthi and

below the right eye, recorded the horizontal and vertical electro-oculograms. Preprocessing was performed with MNE-Python (Gramfort et al., 2014; RRID:SCR_005972) and further data analysis with Fieldtrip (Oostenveld et al., 2011; RRID:SCR_004849) and custom-made MATLAB scripts (The MathWorks).

Data were filtered with a zero-phase bandpass finite impulse response (FIR) filter between 1 and 100 Hz using the window design method ["firwin" in SciPy (<https://docs.scipy.org/doc/>); Hanning window; 1-Hz lower transition bandwidth; 25-Hz upper transition bandwidth; 3.3-s filter length]. A band-stop notch FIR filter from 49 to 51 Hz (6.6 s filter length) was applied to remove line noise. In the next analysis step, data were downsampled to 256 Hz and epoched from −1.5 to 1.5 s relative to the onset of the SIFI-task stimuli. Trials with artefacts (eye blinks, noise or muscle activity) were removed after visual inspection. Data were then re-referenced to the average of all channels and subjected to independent component analysis (ICA) using the extended-infomax algorithm (Lee et al., 1999). Components representing eye blinks, cardiac and muscle activity were removed from the data. Next, noisy channels were rejected after visual inspection on a trial-by-trial basis and interpolated using spherical spline interpolation (Perrin et al., 1989). Finally, trials with signal exceeding $\pm 150 \mu\text{V}$ were excluded. On average, across participants, 105.1 (SD 78.3) trials and 12.7 (SD 4.2) ICA components were removed and 11.4 (SD 2.8) channels were interpolated.

Time-frequency analysis of power

In order to analyze oscillatory power, single trial data were first transformed to time-frequency representations (TFRs). A Hanning taper was applied to an adaptive time window of four cycles for each frequency from 2 to 40 Hz, shifted from −1.5 to 1.5 s, in steps of 10 ms. Poststimulus power was baseline corrected using the average power of the prestimulus window from −500 to −100 ms, relative to the onset of the SIFI-task stimuli.

Analysis of power before the SIFI-task stimuli

To analyze how WM load influences power before the presentation of the SIFI-task stimuli (−1.5 to 0 s), TFRs for the 0-back and 2-back conditions were estimated using trials from all audiovisual conditions of the SIFI task. To minimize the influence of different signal-to-noise ratios and residual motor responses, the number of 0-back and 2-back trials was equalized before the estimation of TFRs. For every trial in the smaller trial set we first selected trials from the larger trial set that matched this trial in terms of the presence of a response to the n -back stimulus in the window preceding the SIFI-task stimulus onset (−1.5 to 0 s). One trial from this subset of selected trials was randomly picked and then removed from the larger trial set before the next trial selection. This selection process was applied to minimize the potential imbalance between conditions regarding the number of trials with n -back task response in the window before SIFI-task stimulus onset.

To assess differences in power between 2-back and 0-back conditions, a nonparametric cluster-based permutation test was conducted (cluster-forming $\alpha = 0.05$, dependent t test, iterations = 1000; Maris and Oostenveld, 2007). The test addresses the multiple comparison problem by clustering samples adjacent in time, frequency, and space. The cluster-based permutation test was applied in the time window from −1.5 to 0 s relative to SIFI-task stimulus onset, on frequencies from 2 to 40 Hz. The observed test statistic was evaluated against the permutation distribution to test the H_0 of no difference between conditions (two-tailed test, $\alpha = 0.025$).

Next, an adjusted version of the cluster-based permutation test was used to investigate the relationships between the differences in power (2-back minus 0-back) and the corresponding change in n -back performance parameters, i.e., sensitivity d' values and RTs (cluster-forming $\alpha = 0.05$, Spearman's rank correlation, iterations = 1000). The correlation analysis was performed separately for each of the two significant power differences derived from the analysis of power in the previous step (Fig. 3). The definition of channels, time intervals and frequency range for the correlation analysis was guided by the characteristics of each cluster. Therefore, the correlation analysis focused on 4–7 Hz and the interval from −1.29 to 0 s for the first cluster (Fig. 3A), and on 20–35 Hz and the interval from −1.47 to 0 s for the second cluster (Fig. 3B). Because of the wide spatial extent of the clusters, the selection of the channels was conservative and included only cluster channels with a total number of significant time-frequency samples at or above the 90th percentile (z score > 1.645). For each cluster the average power across the selected frequencies and channels was estimated at each time point of the cluster-specific interval for both 0-back and 2-back. The cluster-based permutation test was then applied to assess correlations between the load-dependent power difference (2-back minus 0-back) and the corresponding change in the behavioral parameters (Δn -back d' , Δn -back RT). Time clusters derived from the correlation analysis were considered significant only if their p values were below the threshold (two-tailed test, $\alpha = 0.025$). For exploratory purposes, the relationships between the load-dependent power difference (2-back minus 0-back) in significant time clusters and the change in illusion perception in the critical A_2V_1 trials (i.e., Δ illusion rate) was also examined.

Analysis of poststimulus power in the A_2V_1 trials

The analysis of poststimulus power focused on the critical A_2V_1 trials from the SIFI task. The aim was to investigate the effect of WM load on the processing and perception of the SIFI. To this end, a 2×2 repeated-measures ANOVA (Trujillo-Ortiz et al., 2004; https://github.com/juliankeil/VirtualTools/blob/master/vt_freq_bwANOVA.m) on the poststimulus power in A_2V_1 trials was performed with factors Load (low, high) and Perception (no illusion, illusion). The α criterion was set to $p < 0.01$. The analysis included all channels and focused on the time window from 0 to 0.5 s and on frequencies from 2 to 40 Hz. To ensure adequate signal-to-noise ratio in the time-frequency analysis, a minimum number of 30 trials for each condition was required (Luck, 2005). After excluding trials in which no response was provided, an average number of 170.9 (SD 27) artifact-free A_2V_1 trials were available for the 0-back condition and 175.7 (SD 15) trials for the 2-back condition. Hence, a number of 30 trials for each of the four conditions was achieved only when illusion rate in both 0-back and

2-back was approximately between 17.5% and 82.5%. From the 21 participants included in the analysis of power before the SIFI-task stimuli, six participants with an insufficient number of illusion trials (mean illusion rate $\pm$ SD: $12.1 \pm 2.3\%$ in 0-back and $9.9 \pm 5.6\%$ in 2-back) and three participants with an insufficient number of no-illusion trials (mean illusion rate $\pm$ SD: $78 \pm 2.3\%$ in 0-back and $88.3 \pm 2.5\%$ in 2-back) were excluded. Consequently, a subset of 12 participants was included in the further analysis (mean age $\pm$ SD: 26.8 ± 8.3 years; eight females). Moreover, to minimize the influence of different signal-to-noise ratios the number of trials in the four conditions was equalized before TFR calculation. This was done by selecting from the conditions with more trials a reduced set of trials matching as close as possible in RT the trials of the condition with the least trials.

A two-step correction approach was used to address the multiple comparison problem. In the first step, the effects were classified as significant only if at least two neighboring channels showed the same effect (Picton et al., 2000; Maris and Oostenveld, 2007). Then the FieldTrip clustering algorithm (Oostenveld et al., 2011; <https://github.com/fieldtrip/fieldtrip/blob/master/private/findcluster.m>) was applied on the three-dimensional binary matrix that identified the samples meeting the α criterion of $p < 0.01$ (marked as 1; the rest were marked as 0), to create clusters based on temporal, spectral and spatial adjacency. As a second correction step, we used the 3dClustSim algorithm (AFNI, version 17.3.07; Cox, 1996; RRID:SCR_005927) to simulate 10,000 matrices of random values between 0 and 1, with the same dimensions as our data. In these simulations, 3dClustSim estimated the cluster size of connected values below 0.01. Across the simulations, the probability to obtain a three-dimensional cluster of a given size in random data is estimated. Accordingly, we considered clusters significant if they comprised more than 131.7 elements below the α criterion ($p < 0.01$) of the $126 \times 39 \times 51$ (channels $\times$ frequency bins $\times$ time points) matrix. To further analyze main effects and interactions, the analysis was complemented by *post hoc* paired-samples t tests using the Holm–Bonferroni correction (Holm, 1979) to account for multiple comparisons.

Source analysis

The source space analysis was conducted to further investigate the effects obtained from the sensor level analysis. For each participant, the individual T1-weighted MRI (3T Magnetom TIM Trio, Siemens, AG) was co-registered with the individually digitized EEG electrode positions (FastTrak Polhemus) to a common coordinate system (Montreal Neurologic Institute; MNI). This was done by using the digitized head-shape information and the fiducial locations (nasion, left and right preauricular points). The co-registered MRI image was then segmented using the SPM12 algorithm (FieldTrip) and a realistic three-shell (brain, skull, skin) boundary element volume conductor model (BEM) was constructed (Oostendorp and van Oosterom, 1989). Then, the template MNI brain was non-linearly warped onto each participant's anatomic data to obtain a three-dimensional source model (volumetric grid) with a resolution of 10 mm, which was used for the further analysis. To estimate the current density distribution, the eLoreta algorithm (Pascual-Marqui, 2007) was used with a λ regularization parameter set to 1%. The cross-spectral density (CSD) matrix was calculated using the fast Fourier transform (FFT) method for the condition-pooled data in the time interval and the center frequency of each effect, as obtained from the scalp level analysis. Spectral smoothing was defined to fit the frequency of interest (e.g., for the θ prestimulus effect, a center frequency 6 ± 2 Hz smoothing was used resulting in a 4–8 Hz range; Fig. 3A). If short time intervals required extensive smoothing beyond the frequencies of interest, smoothing was defined to the minimum needed for the CSD estimation. The current density estimate for each poststimulus effect was normalized to the source estimate for the baseline window (−0.5 to −0.1 s), and the corresponding frequency range using $\log(\text{Poststimulus/Baseline})$. The log-ratio was used as a form of normalization to correct for possible noise or “center of head” bias, i.e., the fact that source activity is often overestimated in the center of the brain. Thus, using the log-ratio increases the sensitivity of the analysis.

To assess differences in prestimulus source power between 2-back versus 0-back conditions, a one-tailed cluster-based permutation test was used (cluster-forming $\alpha = 0.05$, dependent t test, iterations = 1000, final $\alpha = 0.05$). As described above, the source analysis aimed to further explore the findings of the sensor level analysis. Therefore, the direction of the one-tailed tests was determined by the sensor level results. For each of the poststimulus interaction effects a similar cluster analysis was performed to assess whether source activity related to illusion perception ($\Delta_{\text{illusion-noillusion}}$), supposedly reflecting strong integration, differed between 0-back and 2-back. In line with previous studies (Keil et al., 2014; Balz et al., 2016b), we assume that the difference in neural activity between illusion versus no-illusion trials reveals correlates of integration between the crossmodal signals. The difference $\Delta_{\text{illusion-noillusion}}$ was estimated using $\log(\text{Illusion/Noillusion})$.

Results

Behavior

n-back

Behavioral analysis of the *n*-back task performance (Fig. 2A) revealed that the sensitivity d' values in 2-back trials were significantly lower compared with 0-back trials (mean $\pm$ SD: 3.32 ± 0.64 vs 4.72 ± 0.15 ; $t_{(23)} = -10.4$, $BF = 30,419,046.9$, $p < 0.001$). Furthermore, RTs in 2-back trials were significantly slower compared with 0-back trials (mean $\pm$ SD: 902.3 ± 130 ms vs 663.8 ± 105 ms; $t_{(23)} = 9.9$, $BF = 11,852,368.2$, $p < 0.001$). Hence, higher WM load was associated with worse *n*-back performance.

SIFI

Replicating the finding of our recent behavioral study (Michail and Keil, 2018) in an independent sample ($n = 24$), the SIFI illusion rate was significantly increased in 2-back compared with 0-back trials (mean $\pm$ SD: $39.5 \pm 28.4\%$ vs $35.8 \pm 23.2\%$, respectively; one-tailed paired-samples t test, $t_{(23)} = 2.1$, $BF = 1.3$, $p = 0.025$; Fig. 2B). However, the average RTs to A_2V_1 trials did not significantly differ between 2-back and 0-back trials (mean $\pm$ SD: 787 ± 90 vs 781 ± 93 ms, respectively; $t_{(23)} = 0.6$, $BF = 0.2$, $p = 0.65$). Further analysis of the RTs and the percentage of trials with correct responses in the five control conditions (A_0V_1 , A_0V_2 , A_1V_1 , A_2V_0 , A_2V_2) revealed no significant differences between 2-back and 0-back conditions (all comparisons $p > 0.05$). Thus, WM load specifically affected the SIFI illusion rate, but not the accuracy and RTs in the control conditions.

Correlation between SIFI illusion rate and *n*-back performance

In the next step, Spearman's rank correlation analyses were conducted to investigate whether WM load-dependent changes (2-back minus 0-back) in illusion rates were related to the *n*-back performance across participants. Load-dependent increase in illusion rates correlated negatively with the *n*-back d' reduction ($\rho = -0.49$, $p = 0.047$, $BF = 2.99$; Fig. 2C) but not with the *n*-back RT slowing ($\rho = -0.14$, $p = 0.729$, $BF = 0.20$; Fig. 2C). Hence, participants with higher load-dependent increase in the SIFI illusion rate were less accurate in the *n*-back task (2-back vs 0-back condition).

Neural oscillations

WM load increases θ power and decreases β power before SIFI-task stimuli

The first aim of the analysis was to characterize the effect of WM load on the power in the interval before the presentation of the SIFI-task stimuli. The cluster-based permutation test, performed in the window from -1.5 to 0 s relative to stimulus onset,

Figure 2. Behavioral results of the *n*-back task and the critical A_2V_1 trials of the SIFI task. **A**, Participants showed higher sensitivity d' (left panel) and shorter RTs (right panel) in 0-back compared with 2-back trials. **B**, SIFI illusion rates were higher in the 2-back compared with the 0-back condition (left panel), whereas RTs did not significantly differ between conditions (right panel). Horizontal lines denote the mean and vertical the SEM. **C**, Correlation between load-dependent (2-back minus 0-back) changes in SIFI illusion rates and the corresponding changes in *n*-back d' values (left panel) and *n*-back RTs (right panel). Increased SIFI illusion perception correlated with decreased d' values in the *n*-back task (i.e., worse *n*-back accuracy). Black lines represent the best-fitting linear regression and shaded areas the 95% confidence interval; * $p < 0.05$, *** $p < 0.001$.

revealed significant power differences in the θ and β frequency bands between 2-back and 0-back conditions.

First, power in a frequency range reflecting mainly the θ band (2–9 Hz) was stronger for 2-back compared with the 0-back condition in the interval from -1.29 to 0 s (nonparametric permutation test, $p = 0.0034$). This cluster comprised primarily frontal but also occipital channels (Fig. 3A, left panel). Source analysis of the effect revealed significantly higher θ (4–7 Hz) activity for 2-back compared with the 0-back condition in bilateral prefrontal cortex (PFC) and anterior cingulate cortex (ACC; one-tailed nonparametric permutation test, $p = 0.034$; Fig. 3A, right panel). Second, the 2-back compared with the 0-back condition was associated with lower power in a broad-band frequency range from 6 to 40 Hz, with the strongest effect in the β

Figure 3. WM load-dependent modulation of power before the SIFI task. The cluster analysis revealed two clusters of power difference between 2-back and 0-back conditions. **A**, Frontal θ (4–7 Hz) power, localized in PFC and ACC, was significantly stronger in the 2-back compared with the 0-back condition. This effect was not related to performance changes in the n -back task. **B**, Fronto-central β (20–35 Hz) power, localized in bilateral motor and medial cingulate cortex, was lower in 2-back compared with the 0-back condition. β power decrease was related to load-dependent (2-back minus 0-back) RT slowing in the n -back task. Left panels, TFRs of load-dependent power difference (in t values), averaged across channels with the highest contribution to the cluster and masked based on the temporal and spectral extent of the cluster. Higher values indicate stronger power for the 2-back compared with the 0-back condition. The color scale refers only to unmasked t values. The topographic maps show the spatial distribution of the difference in the cluster's time-frequency window. Channels with high contribution to the cluster (i.e., with a total number of significant time-frequency samples at or above the mean) are highlighted with dots. Middle panels, Time course of the correlation between the load-dependent power difference in the cluster and the corresponding changes in n -back performance parameters, sensitivity $\Delta d'$ (pink) and Δ RT (green). Horizontal lines at the bottom indicate correlation time clusters with $p < 0.1$ and bold letters $p < 0.025$. Right panels, Source contrast (in t values) between 2-back and 0-back for the clusters obtained from the scalp level analysis.

frequency range (~ 20 – 35 Hz; nonparametric permutation test, $p = 0.002$; Fig. 3B, left panel). This cluster comprised fronto-central channels and was observed in the interval from -1.47 to 0 s. Source analysis of this effect revealed significantly lower β power for the 2-back compared with the 0-back condition in a widespread central brain region, including bilateral motor areas and medial cingulate cortex (one-tailed nonparametric permutation test, $p = 0.002$; Fig. 3B, right panel). β power difference (2-back minus 0-back) at scalp level, averaged across frequencies and channels, was then estimated for each time point of the cluster. Next, a cluster-based permutation analysis was used to test whether the β power difference (2-back minus 0-back) correlated at any time point with the load-dependent (2-back minus 0-back) changes in n -back performance. Interestingly, the difference in β power correlated significantly with n -back RT differences in two prestimulus intervals (nonparametric permutation test, $p < 0.025$), but not with sensitivity d' differences (Fig. 3B, middle panel). Load-dependent θ power difference did not correlate with changes in n -back RT or sensitivity d' (Fig. 3A, middle panel). Hence, a stronger load-dependent reduction of β power before the SIFI-task stimuli was related to longer n -back reactions times (2-back minus 0-back). The exploratory analysis of relationships between the average load-dependent power difference in these two intervals and the change in illusion perception in the critical A_2V_1 trials from the SIFI task did not reveal any significant effects (all p s > 0.08). Taken together, our analyses revealed that increased WM load is manifested in increased θ power in bilateral PFC and performance-

relevant modulation in β power in bilateral motor areas and medial cingulate cortex.

Poststimulus θ and β power reflect the interaction between memory load and illusion perception

Poststimulus power was analyzed with a focus on how illusion perception, as well as the different load levels are reflected in the oscillatory power following the presentation of the critical A_2V_1 stimuli. To this end, a 2×2 repeated-measured ANOVA with factors Load (low, high) and Perception (no illusion, illusion) was conducted for the oscillatory power of A_2V_1 trials, in the window from 0 to 0.5 s relative to stimulus onset. The significant main effects and interactions are represented in clusters obtained from the three-dimensional (time-frequency-channel) ANOVA results (for details, see Materials and Methods). For all reported *post hoc* t tests, the relative power change for each condition is provided as mean $\pm$ SD.

The ANOVA revealed a main effect of Load, represented in five channel-time-frequency clusters of oscillatory power. Figure 4A–E illustrates the spectral, temporal, and spatial extent of the effects and the *post hoc* paired-samples t tests comparing the average power of each cluster between 0-back and 2-back. γ (28–40 Hz) power over central-parietal channels, in the interval from 400 to 450 ms, as well as upper θ (7–8 Hz) power over right temporal channels, in the interval from 410 to 500 ms, were significantly higher in 2-back than in 0-back conditions (Fig. 4A,B, respectively). In contrast, low θ power (4–5 Hz) was significantly stronger in 0-back compared with 2-back throughout almost the entire poststimulus interval over frontal and occipital areas (Fig.

Figure 4. Main effect of WM load and perception on poststimulus power. The main effect of Load (low, high) was represented in five clusters of oscillatory power (A–E), and the main effect of Perception (no illusion, illusion) in one cluster (F). Left panel, TFR of each cluster (in F values), averaged over channels contributing to the cluster. Middle panel, Topographic map showing the spatial distribution and the contributing channels (dots). Right panel, *Post hoc* comparisons of the average power of the cluster between the two conditions. Horizontal lines denote the mean and vertical the SEM.

4C–E). There was also a main effect of Perception in the early occipital α band power (7–13 Hz, 0–60 ms; Fig. 4F). *Post hoc* analysis showed that poststimulus α power increase was significantly higher in illusion than no-illusion trials.

More importantly, the ANOVA revealed three clusters of Load $\times$ Perception interactions (Fig. 5). Source reconstruction

was performed for each cluster to assess, at source level, whether activity related to the illusion perception (i.e., $\Delta_{\text{illusion-noillusion}}$, hereafter referred to as illusion-dependent activity) differed between 2-back and 0-back. The first interaction cluster was observed in early β power (18–26 Hz, 0–90 ms) and comprised central-left channels (Fig. 5A). *Post hoc* analysis demonstrated an inverse pattern of early β power modulation between 2-back and 0-back condition (all t tests: $p < 0.05$). In the 0-back condition, β power increase was stronger when participants perceived the illusion ($8 \pm 12\%$) compared with when they did not perceive the illusion ($-12 \pm 9\%$; $t_{(11)} = 4.4$, $BF = 36.9$, $p = 0.004$). On the contrary, in the 2-back condition, a stronger β power decrease was observed when participants perceived the illusion ($-8 \pm 12\%$) compared with when they did not perceive the illusion ($0.6 \pm 16\%$; $t_{(11)} = -2.4$, $BF = 2.2$, $p = 0.035$). Moreover, the relative power change differed between 0-back and 2-back conditions when participants perceived the illusion (0-back: $8 \pm 12\%$ vs 2-back: $-8 \pm 12\%$; $t_{(11)} = -3.6$, $BF = 11.9$, $p = 0.013$) and also when they did not perceive the illusion (0-back: $-12 \pm 9\%$ vs 2-back: $0.6 \pm 16\%$; $t_{(11)} = 2.8$, $BF = 3.8$, $p = 0.035$). Source analysis revealed that early illusion-dependent (illusion minus no-illusion) β activity is higher in 0-back compared with 2-back condition in brain regions including the left motor region [primary motor cortex (M1), supplementary motor area (SMA), and premotor cortex (PMC)] and areas of the temporal cortex [middle temporal gyrus (MTG) and superior temporal gyrus (STG); one-tailed nonparametric permutation test, $p = 0.021$]. Because of the large interval between the β effect (~ 70 ms) and the average RT in A_2V_1 trials (~ 800 ms), and the fact that RTs in these trials were not different between 0-back and 2-back, we consider it unlikely that the early β effect relates to differences in motor preparation.

The second interaction cluster was observed in the θ range (7–9 Hz) over frontal regions in the interval from 30 to 200 ms (Fig. 5B). *Post hoc* analyses demonstrated condition-specific differences between illusion and no-illusion perception. In the 0-back condition, there were no differences in θ power between illusion and no-illusion perception ($t_{(11)} = -1.5$, $BF = 0.7$, $p = 0.198$). On the contrary, in the 2-back condition, a stronger θ power increase was observed when participants perceived the illusion ($34 \pm 40\%$) compared with when they did not perceive the illusion ($13 \pm 35\%$; $t_{(11)} = 3.1$, $BF = 6.3$, $p = 0.028$). Moreover, the relative power change did not differ between 0-back and 2-back conditions when participants perceived the illusion ($t_{(11)} = 1.8$, $BF = 1$, $p = 0.198$), but was different when they did not perceive the illusion (0-back: $33 \pm 44\%$ vs 2-back: $13 \pm 35\%$; $t_{(11)} = -4$, $BF = 22.7$, $p = 0.008$). Source analysis revealed that illusion-dependent θ activity was stronger in 2-back compared with 0-back in middle cingulate cortex (MCC) and ACC (one-tailed nonparametric permutation test, $p = 0.039$).

The third interaction cluster involved frontal β power (13–22 Hz) in the window from 250 to 380 ms poststimulus (Fig. 5C). *Post hoc* analyses demonstrated condition-specific differences between illusion and no-illusion perception. In the 0-back condition, there were no differences in β power between illusion and no-illusion perception ($t_{(11)} = 1.5$, $BF = 0.7$, $p = 0.275$). On the contrary, in the 2-back condition, a stronger β power suppression was observed when participants perceived the illusion ($-29 \pm 22\%$) compared with when they did not perceive the illusion ($-16 \pm 23\%$; $t_{(11)} = -3.1$, $BF = 5.5$, $p = 0.043$). The relative power change neither

Figure 5. Interaction between WM load and perception in the SIFI. Three clusters of interactions between Load (low, high) and Perception (no illusion, illusion) were observed. **A**, The first interaction cluster was found in β power (18–26 Hz; 0–90 ms) over central-left channels. Source analysis identified a corresponding illusion-dependent activity difference in the left motor and auditory cortex. **B**, The second interaction was observed in frontal θ power (7–9 Hz; 30–200 ms). Source analysis identified higher illusion-dependent θ activity in 2-back compared with 0-back, in MCC and ACC. **C**, The third interaction was found in frontal β power (13–22 Hz; 250–380 ms). Source analysis revealed significant differences between 2-back and 0-back conditions in the right PFC, ACC, and bilateral temporal areas. Left panels, TFRs of significant interactions (in F values), averaged across channels contributing to that particular cluster and masked based on the temporal and spectral extent of the cluster, as well as a topography plot showing the spatial distribution and the contributing channels (dots). The color scale refers only to unmasked F values. Middle panels, *Post hoc* paired-samples t tests comparing the average power change of the cluster between the four conditions. Horizontal lines denote the mean and vertical the SEM. Right panels, Source contrast (in t values) for power modulation differences associated with illusion perception ($\Delta_{\text{illusion-noillusion}}$) between 2-back and 0-back, in the cluster's time-frequency window; * $p < 0.05$, ** $p < 0.01$.

differed between 0-back and 2-back conditions when participants perceived the illusion ($t_{(11)} = -1.7$, $BF = 0.9$, $p = 0.275$), nor when they did not perceive the illusion ($t_{(11)} = 1.8$, $BF = 1.1$, $p = 0.275$). Source analysis revealed that late illusion-dependent β activity was lower for 2-back compared with 0-back in right frontal areas (PFC, ACC) and bilateral temporal areas (left posterior STG, right MTG and STG, one-tailed nonparametric permutation test, $p < 0.05$).

Next, a correlation analysis was performed to investigate the relationships between the three clusters of power changes that were associated with illusion perception in the 2-back condition. The analysis revealed that participants with higher illusion-dependent (illusion minus no-illusion) decrease in early β power also showed higher suppression of late frontal β power (Spearman's rank correlation, $\rho = 0.78$, $p = 0.014$, $BF = 16.9$). However, the analysis revealed no significant correlation between the illusion-dependent early β and θ power modulations (Spearman's rank correlation, $\rho = 0.19$, $p = 0.56$, $BF = 0.3$), and between the

frontal θ and late frontal β power modulations (Spearman's rank correlation, $\rho = 0.41$, $p = 0.37$, $BF = 0.5$).

To explore whether power differences in the baseline window might have affected the results of the poststimulus power analysis, an additional analysis was performed without baseline correction. For this exploratory analysis a less conservative significance threshold of $p = 0.025$ was used. The ANOVA for non-baseline-corrected poststimulus data revealed three clusters of interaction effects that were virtually identical in their temporal, spectral and spatial profile with the interaction effects found for baseline-corrected data (Fig. 5). For this reason, we consider it unlikely that differences in baseline power between WM-load conditions account for the observed interaction effects on post-stimulus power.

Taken together, the interaction between Load and Perception indicates a distinction between high and low WM load with regard to the oscillatory responses characterizing the perception of the SIFI illusion. Under low load, the perception of the illusion

was associated with an early left fronto-central β power increase. On the contrary, under high load, the perception of the illusion was characterized by an early left fronto-central β power decrease, followed by a frontal θ power increase around 120 ms and a frontal β power decrease around 350 ms. Notably, the early and the late β power decrease showed a strong positive correlation across participants.

Discussion

In this study, we examined the influence of WM load on perception-related neural oscillations in the SIFI illusion. We found that high compared with low WM load was associated with higher susceptibility to the illusion. Moreover, we observed a modulation of poststimulus power underlying multisensory integration in SIFI, as revealed by the interaction between load and illusion perception at multiple processing stages. Specifically, illusion perception under high compared with low memory load was associated with the engagement of top-down θ and β power. This suggests that crossmodal interactions in the SIFI are sensitive to a load-dependent manipulation of the available cognitive resources.

Replicating our recent work (Michail and Keil, 2018), we found a higher susceptibility to the SIFI under high WM load. Also, across participants, the changes in illusion susceptibility correlated positively with the amount of cognitive resources used by the n -back task. This finding demonstrates that audiovisual integration in the SIFI is sensitive to the amount of available cognitive resources.

Next, we analyzed the power of neural oscillations before the SIFI task to establish that the orthogonal n -back task was efficient in producing power modulations previously associated with WM processes. In agreement with the well-documented role of frontal θ activity in WM (Gevins et al., 1997; Jensen and Tesche, 2002) we found a load-dependent increase in frontal θ power. Moreover, we observed a load-dependent suppression of β power in bilateral motor areas, which is in line with previous reports of frontal β suppression in WM tasks (Brookes et al., 2011; Heinrichs-Graham and Wilson, 2015; Kornblith et al., 2016) and possibly reflects endogenous content rehearsal during WM maintenance (Spitzer and Haegens, 2017). The bilateral distribution of the effect, which was stronger over the right motor cortex, ipsilateral to the response hand, the persistence of the effect up to the SIFI onset and the long interval between the n -back task response and the onset of the SIFI stimulus (at least 1600 ms) argue against attributing this effect to response-related differences in motor activity. Interestingly, the β suppression correlated with the RT slowing in the n -back task, suggesting that β power modulation might reflect the amount of individual cognitive effort (Tallon-Baudry et al., 2004).

We then examined whether WM load affected the oscillatory signatures of illusion perception in the SIFI. Our analysis of post-stimulus power in A₂V₁ trials revealed an interaction between WM load and illusion perception, comprising three distinct effects. The first effect was observed in left frontal β power at ~70 ms, involving left motor areas (PMC and SMA) and the left auditory cortex. Illusion perception in low load was associated with increased early β power, whereas illusion perception in high load was associated with reduced β power. While traditionally being associated with voluntary movement processes, β oscillations in motor cortex have been also implicated in sensory conflict processing (Huang et al., 2014), consistent with evidence on the role of β oscillations in prediction error processing

(Arnal et al., 2011; Arnal and Giraud, 2012). A motor-auditory cortex communication is consistent with the extensive anatomic and functional bidirectional connections between these areas (Zatorre et al., 2007; Rauschecker and Scott, 2009; Nelson et al., 2013; Cheung et al., 2016; Zhang et al., 2016). Therefore, we argue that the observed β power modulation in auditory and motor cortex might correspond to an audiovisual mismatch signal following early crossmodal interactions. Accordingly, the suppression of β power under high load possibly reflects an early mismatch signal. The scarcity of available cognitive resources under high load might prevent the early resolution of the audiovisual perceptual conflict. This notion is consistent with evidence of early β power suppression at left fronto-central channels during early mismatch evaluation of incongruent audiovisual speech stimuli in the McGurk effect (Roa Romero et al., 2015). On the contrary, the enhancement of β power under low load might reflect a signal of perceived audiovisual congruence, a match signal, as a result of strong early crossmodal interactions facilitated by the abundance of cognitive resources. Therefore, our data suggest that the availability of cognitive resources and stimulus congruence play a critical role in defining the nature of early multisensory integration, perhaps through their joint effect on early crossmodal interactions. This notion is consistent with studies demonstrating that the direction of crossmodal interactions, i.e., enhancement or depression, is influenced by attentional resources allocation (Talsma et al., 2007) and stimulus congruence (Calvert et al., 2000). Hence, the load-dependent effect on early β power presumably reflects the modulation of early crossmodal mismatch processing in the SIFI. Future studies are needed to establish whether the assumed alternate representation of both “mismatch” and “match” by β power modulations in an auditory-motor network is a novel phenomenon (Theves et al., 2020).

Following the early interaction effect in β oscillations, we observed that illusion perception under high load was associated with frontal θ power increase around 120 ms poststimulus, localized in the mid-ACC. Interestingly, no such increase was found in the low load condition. Based on evidence of frontal midline θ activity during conflict detection (Hanslmayr et al., 2008; Nigbur et al., 2012; Töllner et al., 2017), exploration uncertainty (Cavanagh et al., 2012), and prediction error processing (Cavanagh et al., 2010), frontal midline θ activity has been proposed as a mechanism underlying the process of cognitive control (Cavanagh and Frank, 2014). A similar role of frontal θ activity in multisensory settings is supported by studies demonstrating frontal θ power modulations in multisensory divided attention (Keller et al., 2017), after spatially incongruent audiovisual stimulation (Cohen and Donner, 2013) and during integration of incongruent audiovisual speech stimuli in the McGurk effect (Keil et al., 2012; Roa Romero et al., 2016; Fernández et al., 2018). Therefore, the midfrontal θ increase during integration of audiovisual SIFI stimuli under high load might correspond to a signal for enhanced need for top-down control in the face of increased perceptual conflict or uncertainty.

In addition to the top-down θ increase, illusion perception under high load was associated with a subsequent frontal β power decrease around 350 ms. Again, no such effect was observed in the low load condition. The localization of this β power effect in the right PFC and ACC and bilateral temporal cortices is suggestive of a top-down frontal modulation of late integrative sensory processing in multisensory processing areas.

There is growing consensus with regard to the role of β oscillations in conveying top-down influences from higher order to low order sensory areas (Buschman and Miller, 2007; Arnal and Giraud, 2012; Bastos et al., 2015; Fries, 2015; Richter et al., 2017). Moreover, STG is a critical brain area for multisensory integration (Calvert et al., 2000; Beauchamp et al., 2004; Balz et al., 2016a). In accordance with these studies, the observed β power suppression during integration under high load might reflect late top-down integration processing in the multisensory association cortex. This proposal is consistent with evidence of frontal areas modulating sensory processing in superior temporal cortex (Sohoglu et al., 2012; Wild et al., 2012). In line with this proposal, integration of audiovisual speech stimuli in the McGurk effect was associated with late frontal β power decrease (Roa Romero et al., 2015). Taken together, the late β power decrease under high load might correspond to enhanced top-down processing of late crossmodal interactions in the SIFI.

In summary, our study reveals that audiovisual integration under high load is associated with early β power suppression, presumably reflecting audiovisual mismatch detection. This is followed by increased top-down frontal θ power signaling the need for increased control, and a subsequent frontal β decrease, presumably reflecting top-down modulation of late integrative processing. Notably, load affected illusion-related oscillations primarily in association areas in temporal cortex, but not in visual cortex. This suggests that the depletion of cognitive resources influences primarily higher-order multisensory processes but does not necessarily influence processing in primary visual areas. Our results are consistent with the proposal that the engagement of top-down processing is required when conflict or competition between the unisensory components of a multisensory stimulus for resources is high (Talsma et al., 2010). The present findings suggest that neural oscillations underlying integrative crossmodal interactions at multiple processing stages dynamically adapt to changing cognitive demands and available resources. Interestingly, audiovisual integration of incongruent speech stimuli in the McGurk effect was associated with analogous neural responses, namely an early and late β power decrease (Roa Romero et al., 2015) and frontal θ power increase (Roa Romero et al., 2016; Fernández et al., 2018). These remarkable similarities between the SIFI-task under high load and the McGurk effect suggest that θ and β power might reflect general integration mechanisms that are recruited when the integration of conflicting audiovisual stimuli requires more processing resources, either because of stimulus complexity (speech vs non-speech) or because of an orthogonal WM load manipulation. Given the behavioral evidence on the effect of perceptual load on the McGurk effect (Alsus et al., 2005, 2007), future studies should investigate to what extent θ and β power are recruited in the McGurk effect under high cognitive load.

References

- Alsus A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol* 15:839–843.
- Alsus A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. *Exp Brain Res* 183:399–404.
- Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16:390–398.
- Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14:797–801.
- Balz J, Keil J, Roa Romero Y, Mekle R, Schubert F, Aydin S, Ittermann B, Gallinat J, Senkowski D (2016a) GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *Neuroimage* 125:724–730.
- Balz J, R Romero Y, K J, Krebber M, Niedeggen M, Gallinat J, Senkowski D (2016b) Beta/gamma oscillations and event-related potentials indicate aberrant multisensory processing in schizophrenia. *Front Psychol* 7:1896.
- Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85:390–401.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Brookes MJ, Wood JR, Stevenson CM, Zumer JM, White TP, Liddle PF, Morris PG (2011) Changes in brain network activity during working memory tasks: a magnetoencephalography study. *Neuroimage* 55:1804–1815.
- Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315:1860–1862.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Cavanagh JF, Frank MJ (2014) Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci* 18:414–421.
- Cavanagh JF, Frank MJ, Klein TJ, Allen JJB (2010) Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage* 49:3198–3209.
- Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ (2012) Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cereb Cortex* 22:2575–2586.
- Cheung C, Hamilton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. *Elife* 5:e12577.
- Cohen MX, Donner TH (2013) Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *J Neurophysiol* 110:2752–2763.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Fernández LM, Torralba M, Soto-Faraco S (2018) Theta oscillations reflect conflict processing in the perception of the McGurk illusion. *Eur J Neurosci* 48:2630–2641.
- Fries P (2015) Rhythms for cognition: communication through coherence. *Neuron* 88:220–235.
- Gevins A, Smith ME, McEvoy L, Yu D (1997) High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb Cortex* 7:374–385.
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86:446–460.
- Haatveit BC, Sundet K, Hugdahl K, Ueland T, Melle I, Andreassen OA (2010) The validity of *d* prime as a working memory index: results from the “Bergen n-back task.” *J Clin Exp Neuropsychol* 32:871–880.
- Hanslmayr S, Pastötter B, Bäuml KH, Gruber S, Wimber M, Klimesch W (2008) The electrophysiological dynamics of interference during the Stroop task. *J Cogn Neurosci* 20:215–225.
- Heinrichs-Graham E, Wilson TW (2015) Spatiotemporal oscillatory dynamics during the encoding and maintenance phases of a visual working memory task. *Cortex J Cortex* 69:121–130.
- Hirst RJ, McGovern DP, Setti A, Shams L, Newell FN (2020) What you see is what you hear: twenty years of research using the sound-induced flash illusion. *Neurosci Biobehav Rev* 118:759–774.
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70.
- Huang S, Rossi S, Hämäläinen M, Ahveninen J (2014) Auditory conflict resolution correlates with medial-lateral frontal theta/alpha phase synchrony. *PLoS One* 9:e110989.
- Jensen O, Tesche CD (2002) Frontal theta activity in humans increases with memory load in a working memory task. *Eur J Neurosci* 15:1395–1399.

- Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. *Cereb Cortex* 15:646–653.
- Kaiser M, Senkowski D, Busch NA, Balz J, Keil J (2019) Single trial prestimulus oscillations predict perception of the sound-induced flash illusion. *Sci Rep* 9:5983.
- Keil J (2020) Double flash illusions: current findings and future directions. *Front Neurosci* 14:298.
- Keil J, Senkowski D (2018) Neural oscillations orchestrate multisensory processing. *Neuroscientist* 24:609–626.
- Keil J, Müller N, Ihssen N, Weisz N (2012) On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb Cortex* 22:221–231.
- Keil J, Müller N, Hartmann T, Weisz N (2014) Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cereb Cortex* 24:1278–1288.
- Keller AS, Payne L, Sekuler R (2017) Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia* 99:48–63.
- Kornblith S, Buschman TJ, Miller EK (2016) Stimulus load and oscillatory activity in higher cortex. *Cereb Cortex* 26:3772–3784.
- Kumar GV, Halder T, Jaiswal AK, Mukherjee A, Roy D, Banerjee A (2016) Large scale functional brain networks underlying temporal integration of audio-visual speech perception: an EEG study. *Front Psychol* 7:1558.
- Lee TW, Girolami M, Sejnowski TJ (1999) Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput* 11:417–441.
- Luck S (2005) Ten simple rules for designing ERP experiments. In: *Event-related potentials: a methods handbook*, pp 17–32. Cambridge: MIT Press.
- Macaluso E, Noppeney U, Talsma D, Vercillo T, Hartcher-O'Brien J, Adam R (2016) The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisens Res* 29:557–583.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748.
- Michail G, Keil J (2018) High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Sci Rep* 8:11530.
- Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci* 27:4120–4131.
- Nelson A, Schneider DM, Takatoh J, Sakurai K, Wang F, Mooney R (2013) A circuit for motor cortical modulation of auditory cortical activity. *J Neurosci* 33:14342–14353.
- Nigbur R, Cohen MX, Ridderinkhof KR, Stürmer B (2012) Theta dynamics reveal domain-specific control over stimulus and response conflict. *J Cogn Neurosci* 24:1264–1274.
- Oostendorp TF, van Oosterom A (1989) Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *IEEE Trans Biomed Eng* 36:382–391.
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
- Pascual-Marqui RD (2007) Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. *arXiv* 0710.3341.
- Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol* 72:184–187.
- Picton TW, Bentin S, Berg P, Donchin E, Hillyard SA, Johnson R, Miller GA, Ritter W, Ruchkin DS, Rugg MD, Taylor MJ (2000) Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37:127–152.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724.
- Richter CG, Thompson WH, Bosman CA, Fries P (2017) Top-down beta enhances bottom-up gamma. *J Neurosci* 37:6698–6711.
- Roa Romero Y, Senkowski D, Keil J (2015) Early and late beta-band power reflect audiovisual perception in the McGurk illusion. *J Neurophysiol* 113:2342–2350.
- Roa Romero Y, Keil J, Balz J, Gallinat J, Senkowski D (2016) Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia. *J Neurophysiol* 116:1396–1407.
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev* 16:225–237.
- Senkowski D, Talsma D, Herrmann CS, Woldorff MG (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res* 166:411–426.
- Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45:561–571.
- Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14:147–152.
- Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. *J Neurosci* 32:8443–8453.
- Spitzer B, Haegens S (2017) Beyond the status quo: a role for beta oscillations in endogenous content (re)activation. *eNeuro* 4. ENEURO.0170-17.2017.
- Tallon-Baudry C, Mandon S, Freiwald WA, Kreiter AK (2004) Oscillatory synchrony in the monkey temporal lobe correlates with performance in a visual short-term memory task. *Cereb Cortex* 14:713–720.
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:679–690.
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410.
- Theves S, Chan JS, Naumer MJ, Kaiser J (2020) Improving audio-visual temporal perception through training enhances beta-band activity. *Neuroimage* 206:116312.
- Töllner T, Wang Y, Makeig S, Müller HJ, Jung TP, Gramann K (2017) Two independent frontal midline theta oscillations during conflict detection and adaptation in a Simon-type manual reaching task. *J Neurosci* 37:2504–2515.
- Trujillo-Ortiz A, Hernandez-Walls R, Trujillo-Perez R (2004) RMAOV2: two-way repeated measures ANOVA. A MATLAB file. MathWorks. Available at <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=5578>.
- van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014) Multisensory integration: flexible use of general operations. *Neuron* 81:1240–1253.
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638–667.
- Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. *J Neurosci* 32:14010–14021.
- Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8:547–558.
- Zhang S, Xu M, Chang WC, Ma C, Hoang Do JP, Jeong D, Lei T, Fan JL, Dan Y (2016) Organization of long-range inputs and outputs of frontal cortex for top-down control. *Nat Neurosci* 19:1733–1742.